The Patent Office

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

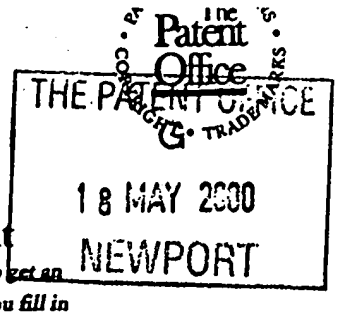
CERTIFIED COPY OF
PRIORITY DOCUMENT

Signed

Dated 23 April 2001

Patents Form 1/77

Patents Act 1977
(Rule 16)



18MAY00 E537650-2 000107
P01/7700 0.00-0011842.2

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

1. Your reference — N00/0385/GB

2. Patent application number *(The Patent Office will fill in this part)* — 0011842.2 — 18 MAY 2000

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

   Power X Limited
   Stafford Court
   145 Washway Road
   Sale
   Cheshire M33 7PE

   Patents ADP number *(if you know it)* — 6803233002

   If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention

   Apparatus and Method for Resource Arbitration

5. Name of your agent *(if you have one)*

   "Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

   McNeight & Lawrence
   Regent House, Heaton Lane
   Stockport, Cheshire SK4 1BS

   Patents ADP number *(if you know it)* — 0001115001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

   a) any applicant named in part 3 is not an inventor, or
   b) there is an inventor who is not named as an applicant, or
   c) any named applicant is a corporate body.
   See note (d))

   Yes

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	7
Claim(s)	
Abstract	
Drawing(s)	5+5/4

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature: McNeight & Lawrence Date: 17 May 2000

12. Name and daytime telephone number of person to contact in the United Kingdom: J G Lawrence 0161 480 6394

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

Apparatus and Method for resource arbitration

TECHNICAL FIELD

The present invention relates generally to systems whose resources include contending requesters and servers, and more particularly to a method and apparatus for arbitrating suitable matches between such requesters and servers.

BACKGROUND OF INVENTION

In many information processing systems having a number of system resources, often called servers, the situation frequently arises wherein a set of requesters must arbitrate for service from one of a number of servers. For instance, in a data packet switching application, the requesters may each need to transmit one or more cells (i.e., small packets of information) to various ones of the servers. Each requester receives cells from one or more system users (e.g., components) that need to be forwarded to specified one of the servers. If the requester has at least one cell that needs to be forwarded to a server, then the requester will generate a request for access. The problem is to rapidly produce a set of one-to-one (point to point) matchings that maximizes the number of connections between requesters and servers. Preferably, the matchings should be produced without imposing significant overheads on the system.

Although the subsequent material herein describes a local area-switching network, the present invention may be applied in any application in which a rapid matching of requestors and servers is needed.

The present invention has application in an arbitration mechanism and procedure that provides maximal matching within a relatively small number of clock cycles. Fast matching is important so that the arbitration procedure can be repeated quickly for successive sets of requests. In particular, the arbitration procedure for one set of requests should be no longer than the amount of time needed to transmit the requests granted during the previous arbitration cycle. When the requests are small cells that are transmitted in just a few clock cycles, the arbitration mechanism has only a similar number of clock cycles to arbitrate competing requests. The arbitration mechanism of the present invention can be implemented using presently existing circuit components.

In one application of the invention each input port receives, and stores in an input buffer, a set of cells, with each cell member bearing an address that identifies an output port to which the cell requests routing. The arbitration mechanism now must satisfy not one request, but a set of requests. This is a relatively difficult and time-consuming task because many arbitration matches are possible between cells in the input port sets and the output ports. Some matches will be better than others will; the problem is to rapidly find a good match satisfying the largest number of contending requests.

Crossbar Arbitration

Switching (i.e., connecting requesters to servers, or transmitting data from senders to receivers) involves two separate tasks. First, a scheduling task is performed, wherein an arbitration mechanism selects which of potentially many requests to send to each server. Second, a data-forwarding task is performed, in which a switching mechanism forwards requests or cells to servers in accordance with the selections made by the scheduling task.

Bipartite Graph Matching

At the center of high-speed communication switches, there is generally a crossbar matrix switch to do the actual switching at the physical layer. Crossbar switches enable point-to-point connections to be configured to connect one switch port to another. The crossbar has to have connections formed across it by some form of algorithm. No two inputs can connect to the same output (fig: 1①) and no two outputs can receive different cells from the same input (fig: 1②), the acceptable solutions are fig: 1③④. This set of connections must be produced from a set of requests that come from the input ports. This is commonly known as a Bipartite Graph Matching problem. There are many different solutions to this problem. Each solution addresses a different weakness and attempts to resolve it.

Pointers

Nearly all Bipartite Graph Matching algorithms involve some mention of pointers, so some explanation is needed. Each input and output port has a pointer that indicates where the arbiter will attempt to construct connections. In most algorithms there is both grant and an accept pointer, these are associated with output and input ports respectively. It is with the manipulation of these pointers that most of the algorithms are concerned.

Fairness

All Bipartite Graph Matching algorithms attempt to allocate potential connections fairly. In practice, this means that over a statistically significant period of time that potential connections will be distributed evenly between all of the ports with connection requests.

Controllability

The definition of fairness used up to this point has been that bandwidth should be allocated as *evenly* as possible between the ports requesting connections. This definition is overly simplistic; in the real world, this definition becomes much too rigid. What is really needed is the ability to *control* what bandwidth is allocated where and to whom. It was with this aim in mind that Probabilistic Masking was developed (GB.9828144.76-22-12-98. Data Switching Apparatus) where connection bandwidth allocation is performed by a Probabilistic Masking unit placed before the arbitration stage. Probabilistic Masking functions by (Pseudo) randomly removing connection requests to a varying degree before they reach the arbitration stage, thus bandwidth on that particular connection is limited enforcing varying levels of Quality of Service. The Probabilistic Masking arrangement

has a limitation that it does not operate in a work conserving manner, when a request is masked potential bandwidth is lost, this bandwidth is not then allocated elsewhere.

It is an object of the invention to provide an arbitration arrangement that overcomes this limitation.

SUMMARY OF THE INVENTION

According to the invention there is provided an arbitration arrangement for use in selecting the connections to be made between ingress and egress ports of a memoryless cross-bar switch of a data switching system, the arbitration arrangement comprising a three phase process involving (i) a request phase in which each ingress port sends its connection requests to the egress ports to which a connection is required, (ii) a grant phase in which each egress port examines the requests directed to it and selects one request for grant returning a grant signal indicative of the selected request and directed to the appropriate ingress port and (iii) an accept phase in which each ingress port examines the received grant signals and selects one to accept thereby defining an ingress to egress port connection across the memoryless cross-bar switch, characterised in that each ingress port is allocated a default weight which is decremented by one each time a connection is made in the cross-bar switch involving that ingress port and a connection request is only generated in the request phase if the value of the weight is more than zero and the weights are reset to their default value when all requests to an egress port have had their weight values decremented to zero.

According to a feature of the invention there is provided an arbitration arrangement in which each ingress port request is applied to a masking element comprising a default weight register, programmed with the default weight to be allocated to the said ingress port by a system control device, a working register whose contents are decremented by one each time a cross-bar switch connection is made involving the said ingress port, the contents of the working register being used to mask the connection request generated by the ingress port when said contents are zero.

Also according to the invention there is provided an arbitration process for use in selecting the connections to be made between ingress and egress ports of a memoryless cross-bar switch of a data switching system the arbitration process comprising a three phase process involving (i) a request phase in which each ingress port sends its connection requests to the egress ports to which a connection is required, (ii) a grant phase in which each egress port examines the requests directed to it and selects one request for grant returning a grant signal indicative of the selected request and directed to the appropriate ingress port and (iii) an accept phase in which each ingress port examines the received grant signals and selects one to accept thereby defining an ingress to egress port connection across the memoryless cross-bar switch, characterised in that each ingress port is allocated a default weight which is decremented by one each time a connection is made in the cross-bar switch involving that ingress port and a connection request is only generated in the request phase if the value of the weight is more than zero and the weights are reset to their default value when all requests to an egress port have had their weight values decremented to zero.

Functional Description

The new arbitration algorithm's function is to take a data set of connection request bits and form connections obeying the rules described above in the crossbar arbitration section. When viewed over a statistically significant amount of time the connections formed by the algorithm are made more or less likely to be formed depending on the weight allocated to that particular link. The higher the weight for a connection the more likely it is for that particular connection to be formed. This ability to define the likelihood of connections being made enables deterministic work-conserving bandwidth management.

Bandwidth Management

The new algorithm incorporates all of the functionality that is provided by a Probabilistic Masking unit into the arbitration stage itself. In addition, to avoid all of the problems associated with probabilistic operation the new arbitration stage behaves in a deterministic way. This is achieved by taking an arbitration stage and adding extra functionality in the form of weights for each connection. When a connection is made, the weight is decremented, if there is no remaining weight the connection cannot be made. This extension gives all of the traffic shaping ability of probabilistic masking, while maintaining the work conserving nature of the arbitration stage.

Deterministic Work Conserving Operation

The masking effect of the weights is similar to that of Probabilistic Masking; however, the new algorithm behaves in a deterministic way. In addition, any residual bandwidth that is not being used is allocated in a work-conserving manner. Bandwidth Allocation guarantees that a minimum level of service will be delivered under high load. Work conserving operation allows the switch to reduce latency under lower loads.

Variable Granularity

The granularity of bandwidth allocation is controlled by the range of integer values that the weighting system can hold, the larger the values the finer the granularity. As the only limitation on the weight values is the size of the register that is used to hold their values, hence, the granularity of bandwidth is easily altered by simply using more bits.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Legal arbitration states. This shows the different possible connection combinations for a two-port switch.

Figure 2: Weight Reset Conditions. This gives an example of weight/request masking.

Figure 3: Single output port request masking process. This details the entire arbitration process for one port.

Figure 4: shows an example of a masking element associated with a request signal.

Figure 5: shows weight reset arrangements for one egress port.

DETAILED DESCRIPTION OF THE PREFERRED ALGORITHMIC EMBODIMENT

The current invention is based upon a three-stage arbitration algorithm. The first and third stages remain essentially the same; the second or grant stage is altered to incorporate the connection bandwidth allocation functionality.

Stage 1 **Request**

For a thirty-two-port system, there are one thousand and twenty four request signals (thirty-two squared); this allows requests for connections from any point to any other point to be requested. Creation of requests is an external matter and is handled by another mechanism.

Stage 2 **Grant**

For each port, two processes are performed to establish which request to grant. Before the round robin process can be used to establish a grant, the weight table for each port must be checked to see if it should be reset to its original allocation values. The reset condition is that *every port with a request has a weight of zero* (figure 2). The lack of a request from a particular input port effectively masks the weight for that port, removing it from the reset calculation. A reset consists of every weight for a particular port being reloaded with its original value. These values are held in another set of registers/RAM that can be written to, to alter the bandwidth allocation parameters of the arbiter.

The second process is a Round Robin with an extra grant condition that is dependent upon the weights for the connection. As the pointer moves around the set of requests, for a request to be granted there must also be enough weight for that particular connection i.e. a value greater than zero. The weights in effect temporarily mask certain requests. Although requests are being rejected, this algorithm remains work conserving by ignoring weights that have no requests relating to them. If there are only request with no weight the weights will be reset (see above) and connections will be made. This work conserving nature guarantees a minimum bandwidth at high load, but allocates more if it becomes available.

The operation of this masking effect is shown in figure three, a single output port from a four-port arbiter. The requests from each input port ① are shown, these are then individually ANDed with the weights ② for each input port. The result ③ is then presented to a RR pointer for a request to be granted. It can be seen that the pointer moves round from the last location ④ to the next location with a valid request ⑤.

Stage 3 **Accept**

In the third stage of the arbitration, each port that has received a grant ⑥ uses a round robin ⑦ to pick which one to accept ⑧. Once an input and an output port have negotiated a connection, they then set the appropriate blocked flags, eliminating themselves from the selection process. Once a connection has been confirmed, the weight for that connection is decremented by one ⑨. The grant and accept pointer values for the connection concerned are then updated ⑩.

The masking effect of the weights is similar to that of Probabilistic Masking; however, the new algorithm behaves in a deterministic way.

DETAILED DESCRIPTION OF THE PREFERRED REGISTER LEVEL EMBODIMENT

Changes to the second part of the arbitration process detailed above are implemented in the form of a masking element that is inserted in front of the grant round robin stages. In a thirty-two-port arbitration stage, there would need to be one thousand and twenty four masking elements. Figure 4 shows the Register Transfer Level (RTL) design of one individual masking element. Figure 5 shows the construction of one egress port from thirty-two masking elements.

Masking Element Operation

This sequence references Figure 4.

1. The weighting for the one request is loaded into the default weight register (1).

First cycle:

2. This default weight held by (1) is then transferred onto the working register (3), via the multiplexor (2), on the clock PGT that signifies the start of the first cycle.

3. The non-zero status of the weight in the working register (3) is determined by the OR gate (4).

4. The status determined by (4) is then used to mask out the request if the status of the working register (3) is zero. This is performed by the AND gate (5).

5. The value of the weight in working register (3) is passed to the Decrementer (8) where if a connection has been received due to a request made on the previous cycle the value of the weight is decremented by one and returned to working register (3).

6. The non-zero status of the decremented weight is determined by the OR gate (6).

7. If there is no request in this cycle and the weight value has been decremented to produce a non–zero result as determined by (6) then the reset condition for this request has been met. This product function is provided by (7).

8. When this condition has been met for all requests to a single Egress port, the default weights for all requests to this Egress port are required to be reloaded into their relevant working registers (3). This is determined by (9), which has a similar connection to all other masking elements for this Egress port.

DIAGRAMS

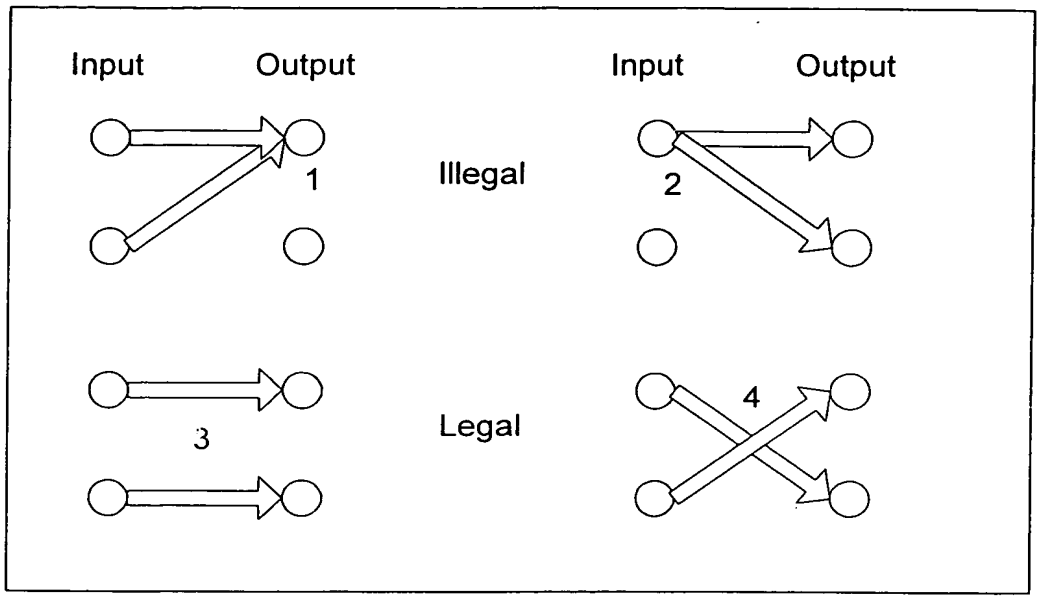


Figure 1: Legal/Illegal arbitration states

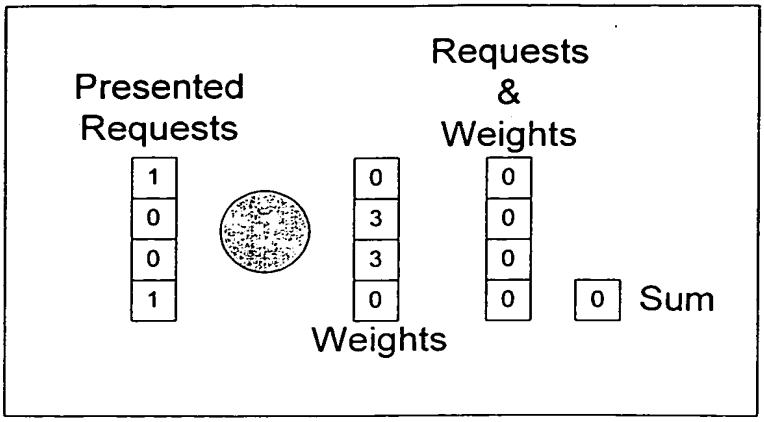


Figure 2: Weight Reset Conditions.

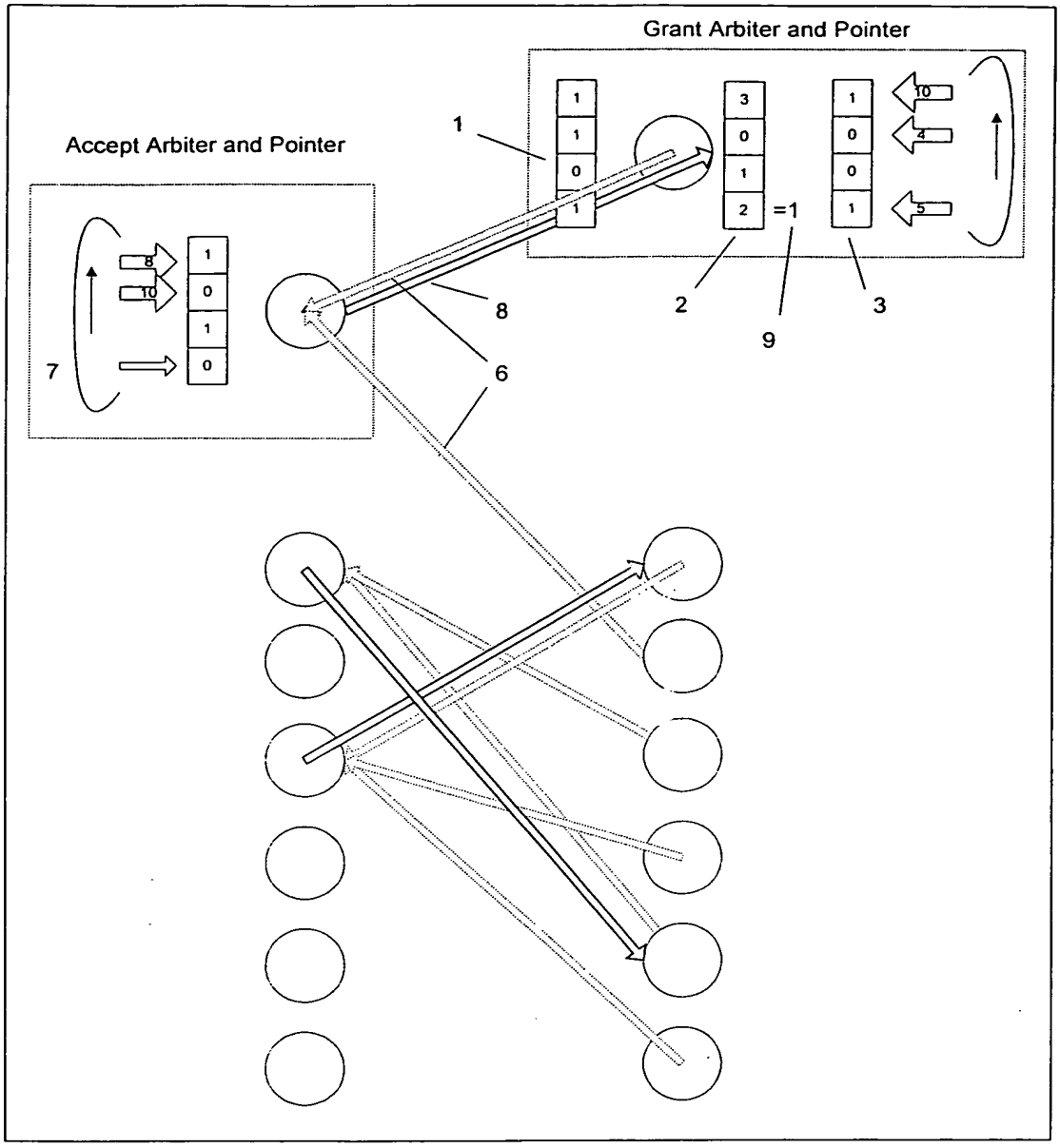


Figure 3: Single output port request masking process.

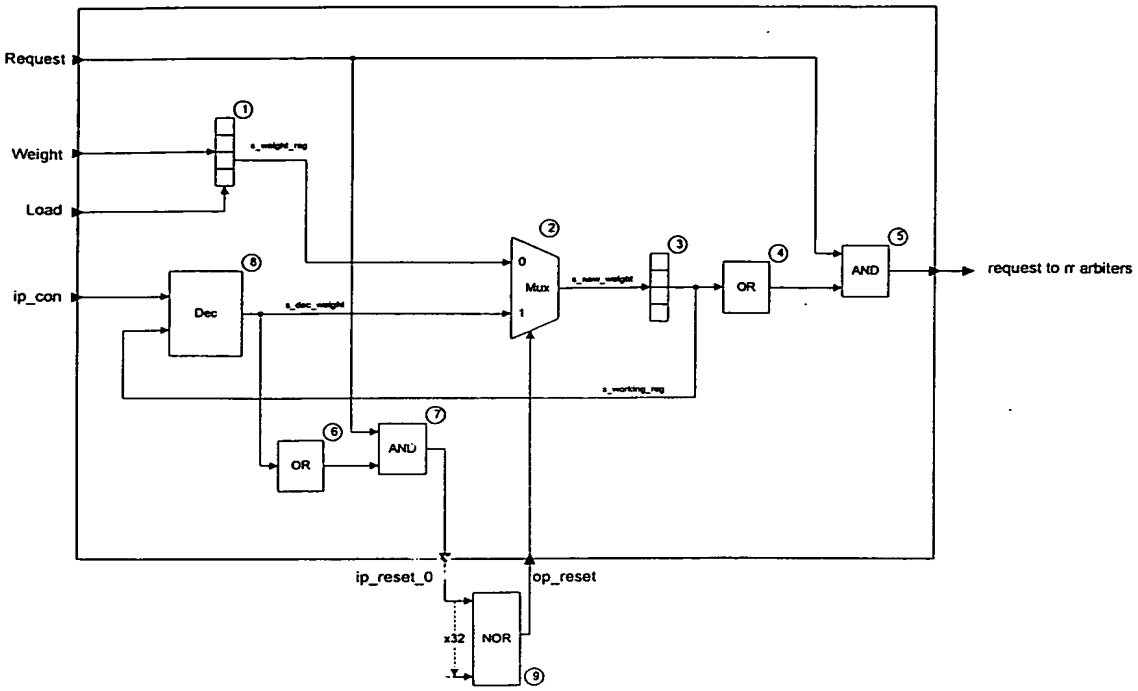


Figure 4: Single Masking Element

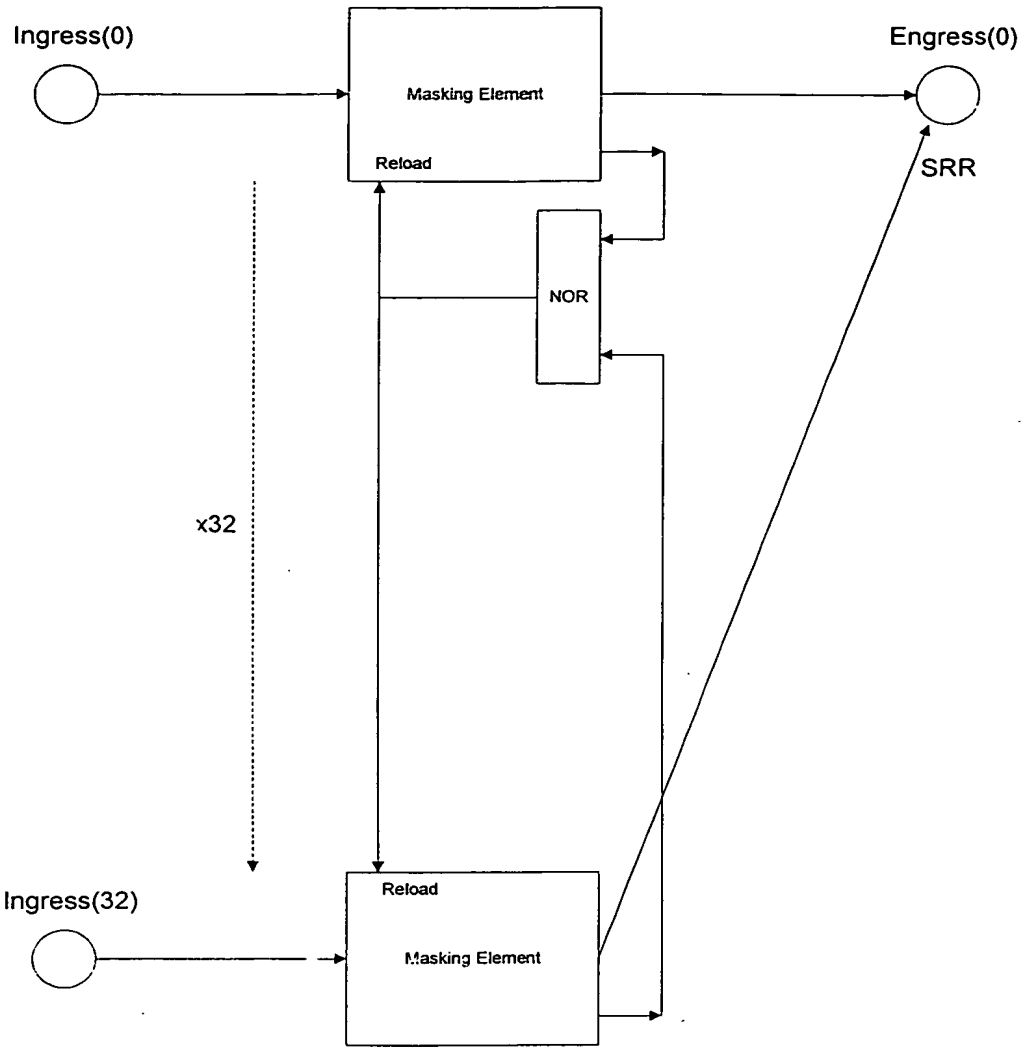


Figure 5: Single Egress Port with Masking Elements